François Rebaudo

R pour les scientifiques

Mise en oeuvre de projets et valorisation des résultats

DUNOD

Illustration de couverture : Login – AdobeStock.com

11 rue Paul Bert, 92240 Malakoff
www.dunod.com
ISBN 978-2-10-081547-0

Table des matières

Avant-propos

Sommaire

0.1 Pourquoi se former à R

Le logiciel et langage de programmation R s'est imposé comme un outil incontournable d'analyse et de gestion des données scientifiques (et des données d'une manière générale). Il devient indispensable dans ce contexte d'en maîtriser *a minima* les bases. Le succès de R n'est pas un hasard : R est un logiciel que tout le monde peut se procurer librement, et son fonctionnement à base de scripts assure la **transparence** et la **reproductibilité** des résultats scientifiques (sous réserve de respecter quelques règles que nous aborderons dans ce livre). R repose sur une **communauté** très active à l'origine de la création de plusieurs milliers de modules complémentaires (packages) permettant les analyses statistiques les plus pointues. Il est disponible sur les principaux systèmes d'exploitation (Linux, OSX et Windows). Les codes R (ou scripts) sont, sauf cas exceptionnels, indépendants du système d'exploitation utilisé, assurant ainsi leur **portabilité**. R est également un outil qui s'adapte à tous les besoins, depuis de simples statistiques descriptives pour des petits jeux de données jusqu'à la gestion et l'analyse de gros de jeux de données (SIG, génomes, BigData …), que ce soit en local sur un ordinateur, ou à distance sur des serveurs.

0.2 Les objectifs de l'ouvrage

Ce livre est né de la demande des étudiants et professionnels que j'ai eu la chance de rencontrer et de former à R. Les scientifiques, les étudiants et autres personnes souhaitant s'initier à R y trouveront toutes les ressources nécessaires à la mise en œuvre de leurs propres projets scientifiques et à la valorisation de leurs résultats. Il existe de nombreux livres dédiés à R, mais aucun ne couvre les éléments de base de ce langage qui permettent de rendre des résultats scientifiques publiables et reproductibles. Ainsi, tout au long de cet ouvrage nous nous efforcerons de rédiger non seulement un code fonctionnel pour la machine, mais aussi un code lisible et réutilisable pour les humains.

Tous les codes présentés dans ce livre sont disponibles sur dunod.com [1].

0.3 Lectures complémentaires

— “R pour les débutants”, Emmanuel Paradis (`https://cran.r-project.org/doc/contrib/Paradis-rdebuts_fr.pdf`)
— “Introduction à la programmation avec R”, Vincent Goulet (`https://cran.r-project.org/doc/contrib/Goulet_introduction_programmation_R.pdf`)
— Le blog `http://rzine.fr`, un site collaboratif et interdisciplinaire sur la pratique de R en sciences humaines et sociales

0.4 Remerciements

Je remercie toutes celles et ceux qui ont participé à améliorer ce livre par leurs conseils, leurs suggestions de modifications et leurs corrections (par ordre alphabétique) : Camila BF, Marc G, Susi LH, Emmanuel P, Estefania QH, Baptiste R et Jean-Christophe S.

1. `https://www.dunod.com/`

Première partie

Les concepts de base

Chapitre 1

Premiers pas

Sommaire

1.1 L'installation de R

Le logiciel R peut être téléchargé depuis de nombreux serveurs du CRAN (Comprehensive R Archive Network) à travers le monde. Ces serveurs s'appellent des "miroirs". Le choix du miroir est manuel.

Le programme permettant l'installation du logiciel R peut être téléchargé depuis le site web de R : `https://www.r-project.org/`. Sur ce site, il faut au préalable choisir un "miroir CRAN" (serveur depuis lequel télécharger R ; sauf cas particulier, le plus proche de sa localisation géographique), puis télécharger le fichier **base** correspondant à son système d'exploitation. Les utilisateurs de Linux pourront préférer un `sudo apt-get install r-base` (ou équivalent). Au moment d'écrire ce livre, la version disponible est la 4.0.3 dénommée "Bunny-Wunnies Freak Out".

1.2 R comme calculatrice

Une fois le programme lancé, une fenêtre apparaît dont l'aspect peut varier en fonction de votre système d'exploitation (Figure 1.1). Cette fenêtre est dénommée la **console**. La première information que l'on peut trouver sur la console est la version de R utilisée. Il est recommandé de mettre à jour régulièrement sa version de R afin de bénéficier des dernières fonctionnalités.

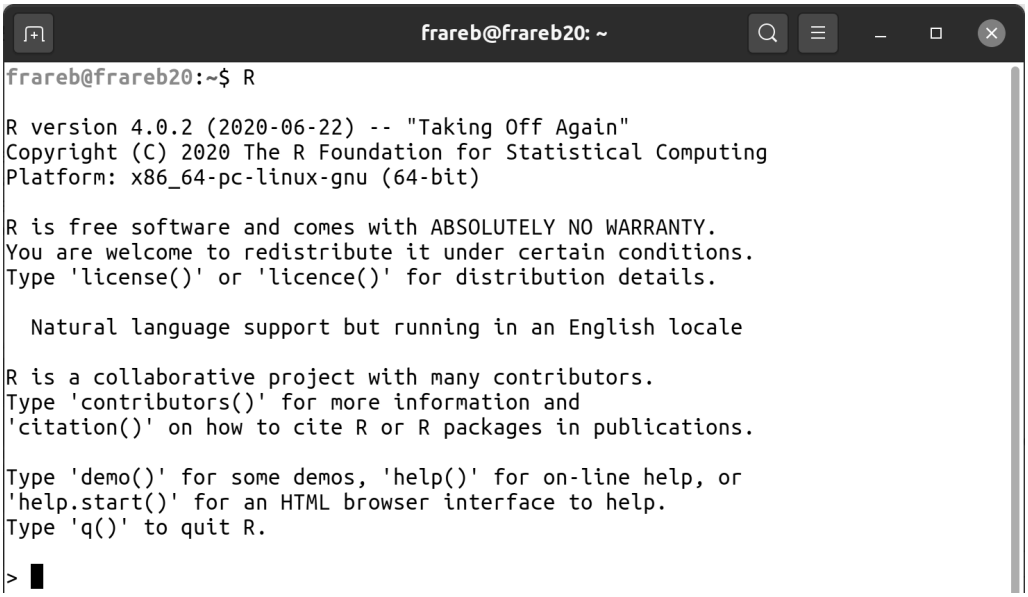

Figure 1.1 – Capture d'écran de la console R sous Linux Ubuntu.

La console correspond à l'interface sur laquelle va être interprété le code, c'est-à-dire l'endroit depuis lequel le code va être transformé en langage machine, puis exécuté par l'ordinateur. Le résultat de cette exécution sera retransmis dans la console sous une forme lisible par des humains. Cela correspond à l'écran d'affichage d'une calculatrice. C'est de cette manière que R va être utilisé dans la suite de cette section.

Tout au long de ce livre, les exemples de code R apparaîtront sur un fond gris. Ils peuvent être copiés et collés directement dans la console (pour la version électronique de ce livre, et alternativement depuis le site dunod.com [1]), bien qu'il soit préférable de reproduire soi-même les exemples dans la console (ou plus tard dans les scripts). Le résultat de ce qui est envoyé à la console apparaîtra sur fond blanc avec "##" précédent le code afin de distinguer le code et le résultat du code.

1.2.1 Les opérateurs arithmétiques

```
5 + 5
```

1. https://www.dunod.com/